Selected Data Manipulation Techniques in R

Using R Studio

Imported Your Data Already?

- If you already have your data in RStudio, you can skip the slides providing an overview of the import process.

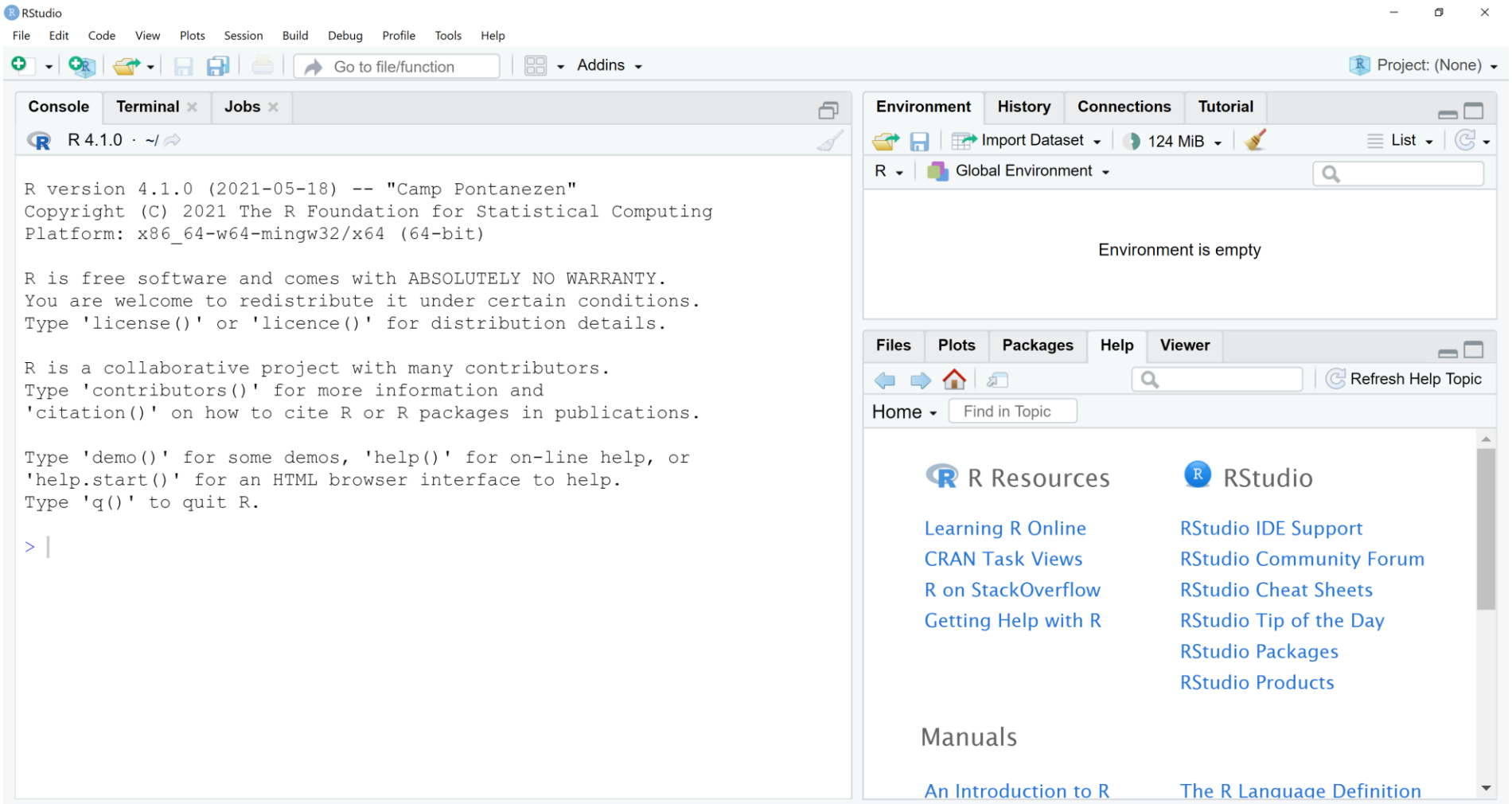
Dataset

- This tutorial is a walkthrough with a sample set of data. You may use this to walk through the tutorial, if you wish, but for your assignments, you will be asked to use your own dataset (as specified within the course).

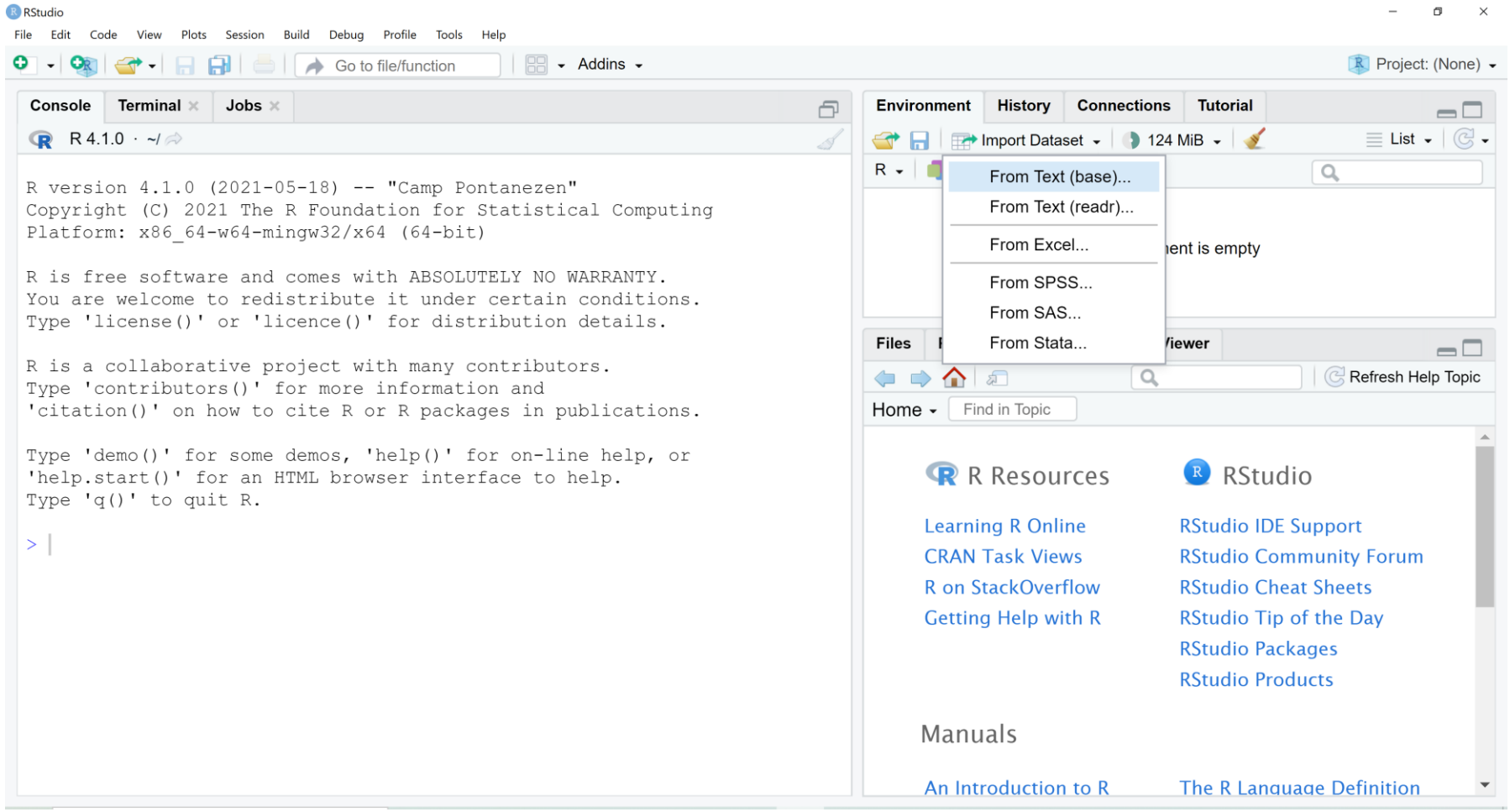
Dataset reference:

Skoryk, M. (2021). Sepsis Prediction from Clinical Data. Version 1. Retrieved from https://www.kaggle.com/maxskoryk/datasepsis

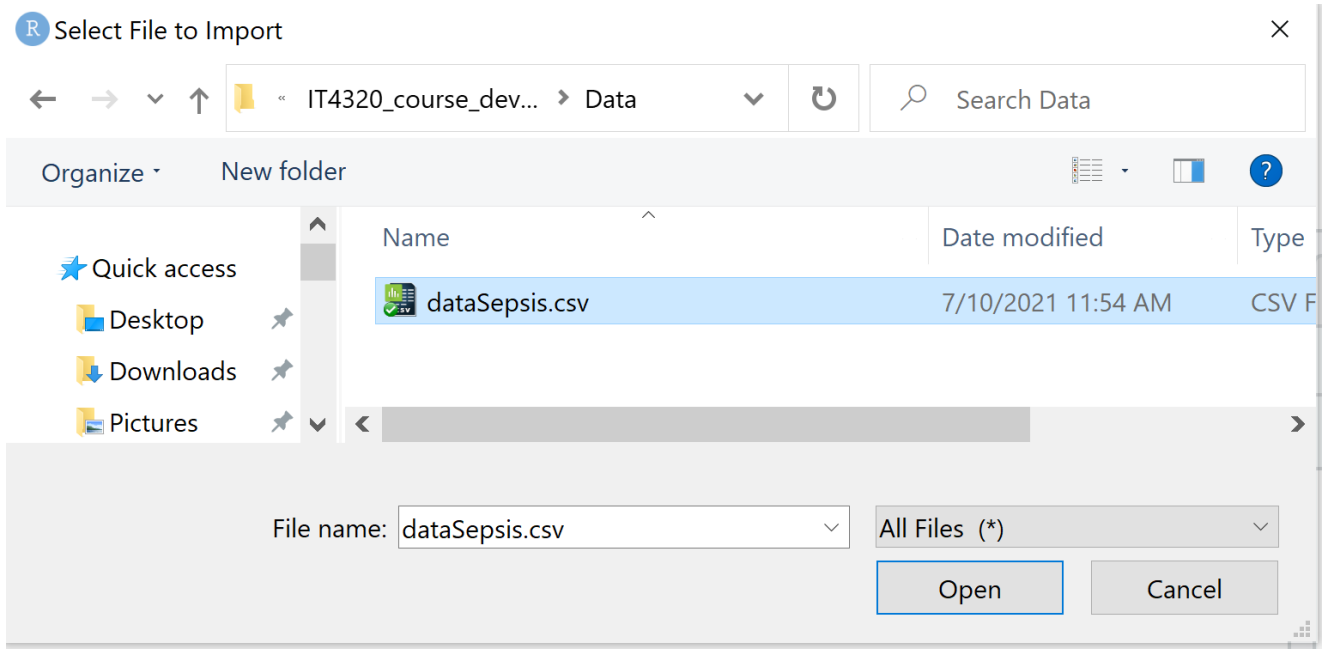
Open R Studio



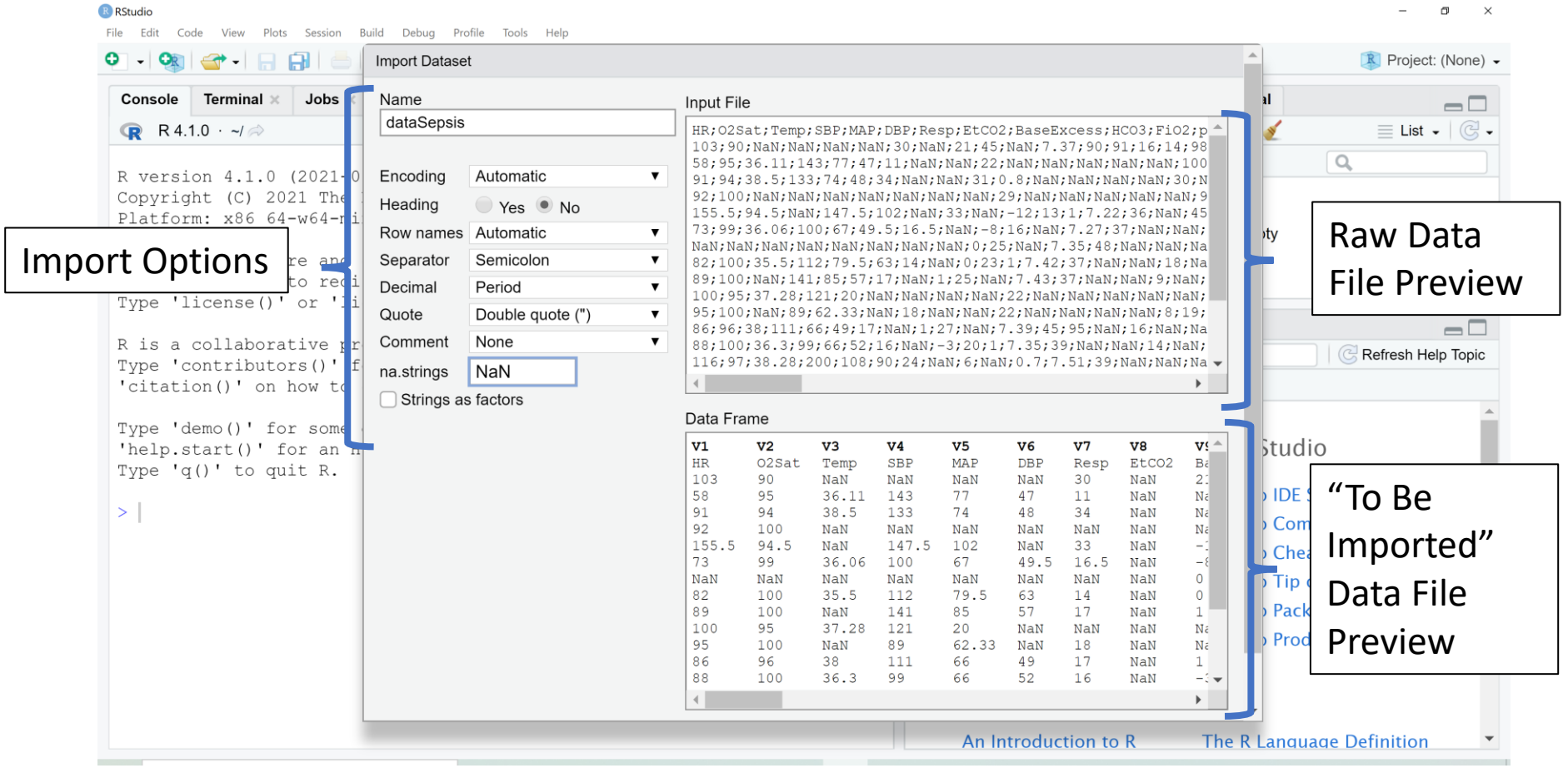
Click on “Import Data” and Choose “From Text (base)”



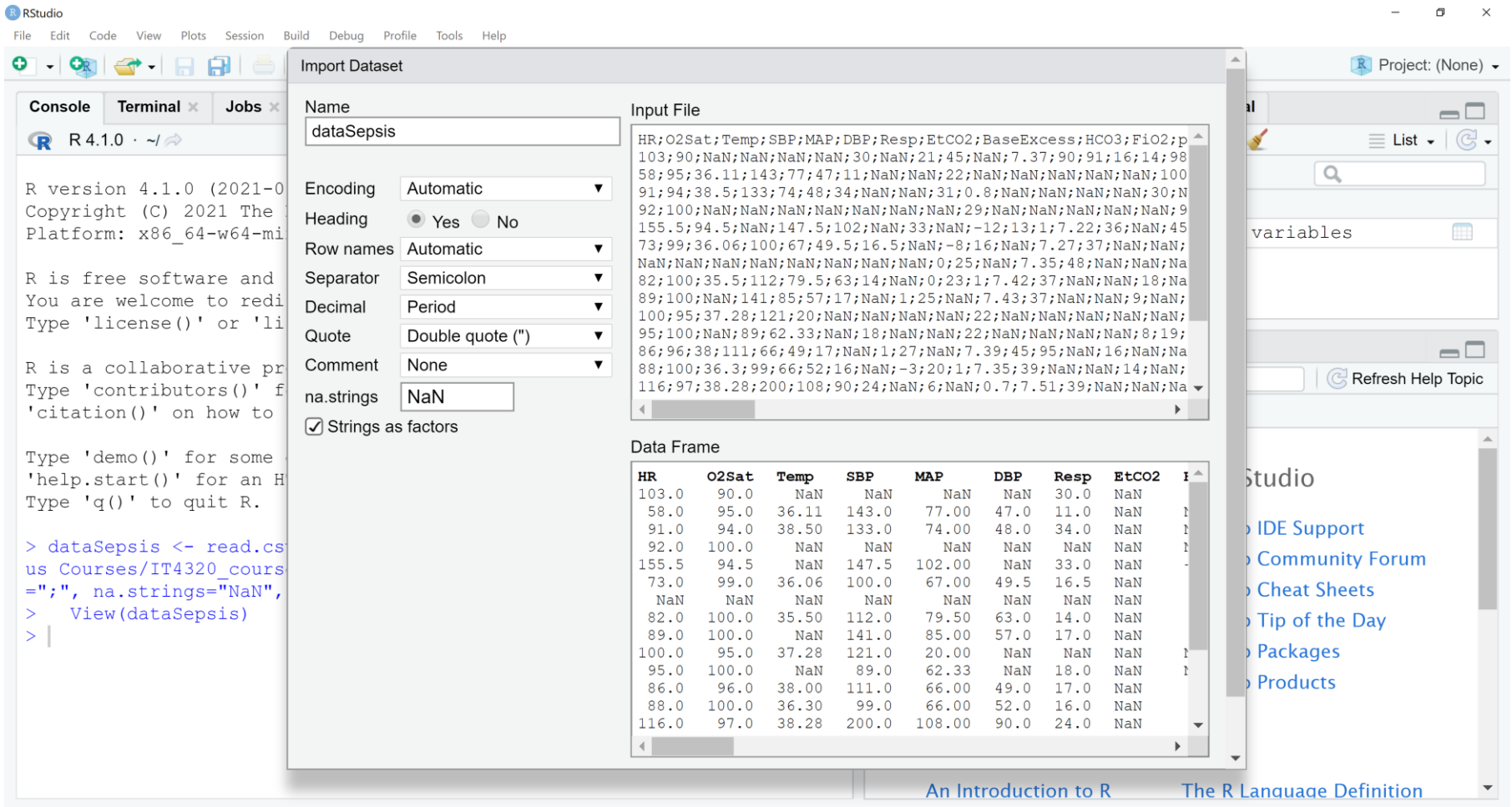
Navigate to Your Dataset, Then Click “Open”



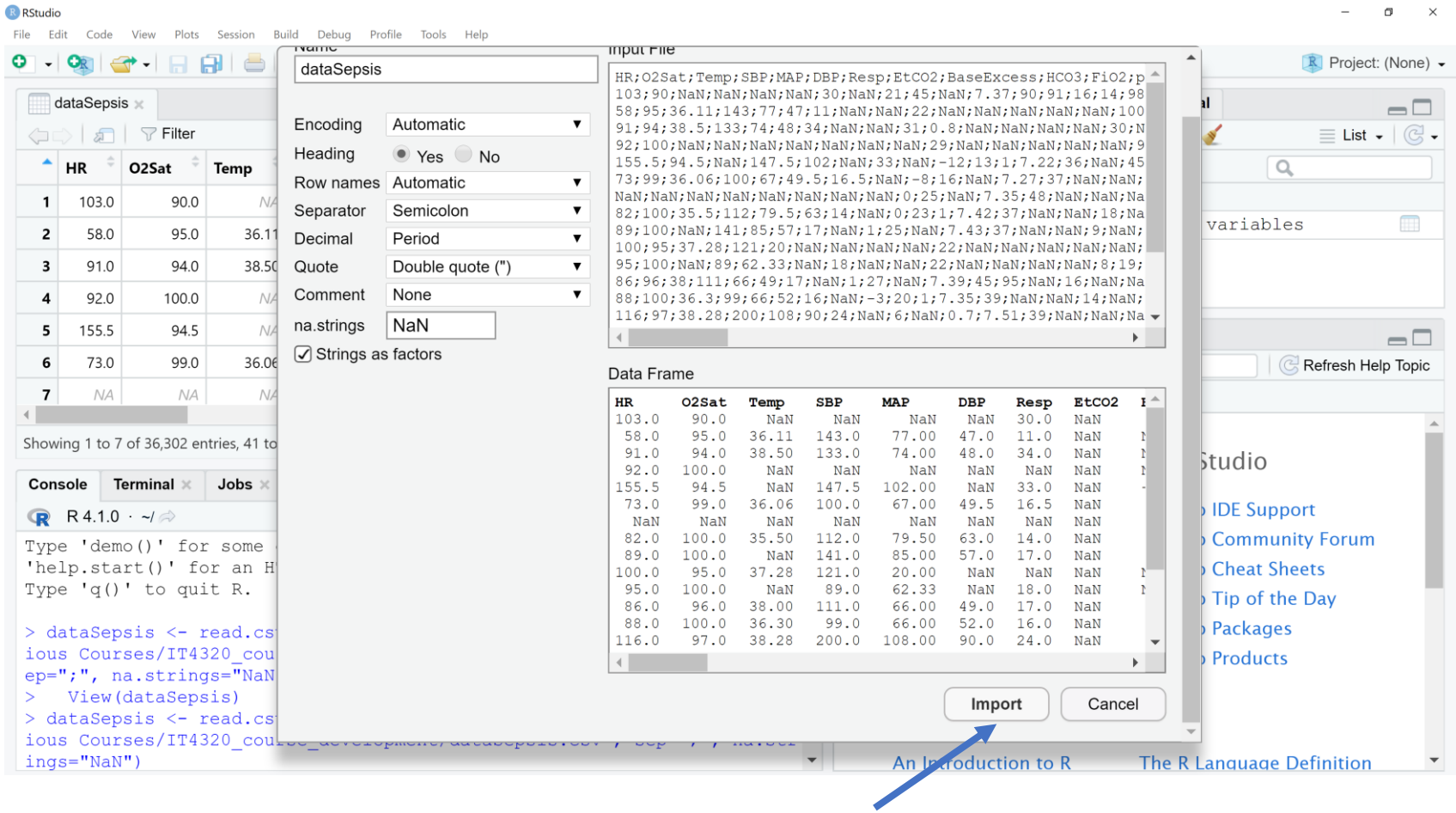
Select Options to Import Your Data Based on the Format of Your Text File



The Options Displayed are Those Required to Successfully Import the sepsis dataset.



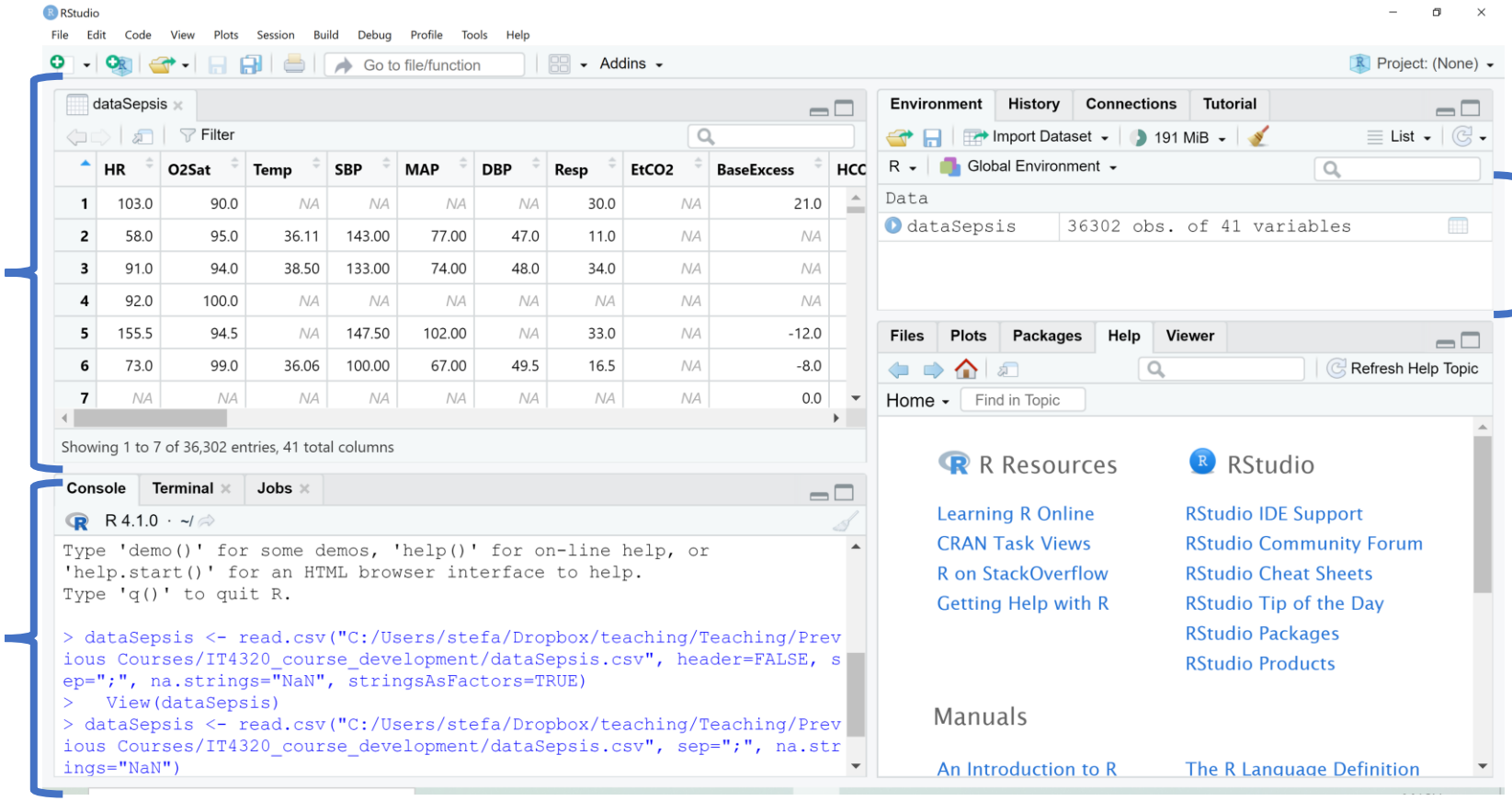
Scroll Down and Click “Import” to Complete Import Process



You May Verify Successful Upload On the Following Screen

Imported Data

Logfile Indicating Options in the Import Process



Dataset Summary

More Options for Importing Data Into R Studio

https://support.rstudio.com/hc/en-us/articles/218611977-Importing-Data-with-the-RStudio-IDE

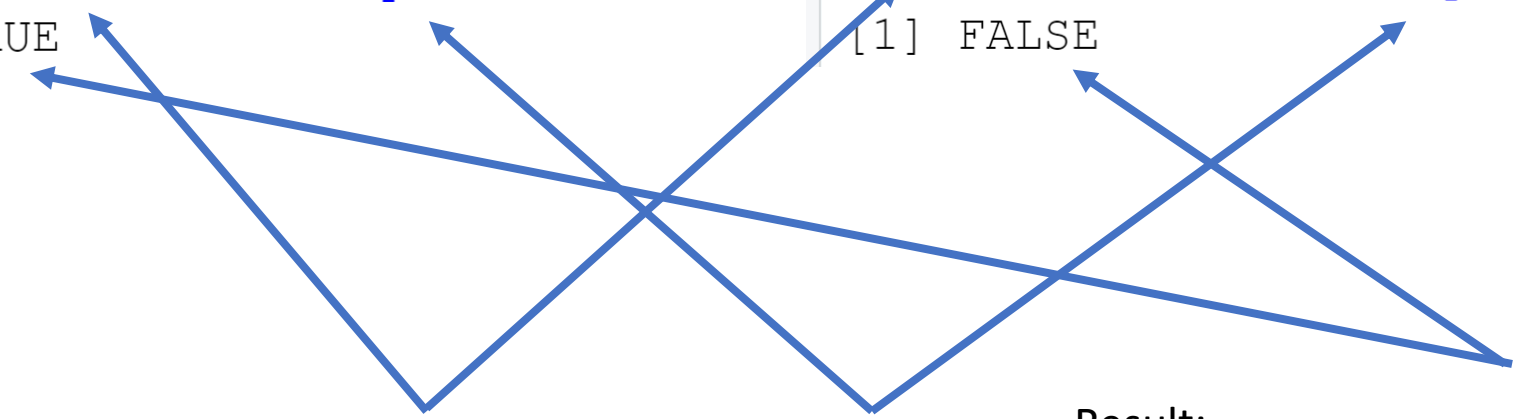
Selected Data Manipulation Examples

- Checking the format of a variable
- Converting a numeric variable to a character variable
- Converting a character variable to a numeric variable
- Creating a calculated variable
- Populating values of a variable using an if-then statement
- Combining multiple character variables
- Taking only a portion of a character variable using a substring

Checking the format of a variable

```
> is.numeric(dataSepsis$HR)
[1] TRUE
```

```
> is.character(dataSepsis$O2Sat)
[1] FALSE
```



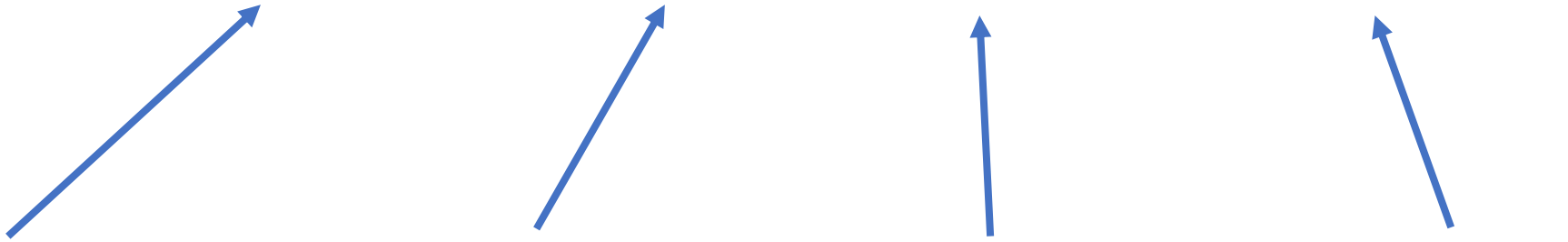
format check function:
Asks R if a variable is a specific format (numeric or character)

Variable to be checked

Result:
TRUE= the variable is the variable type specified in the format check function.
FALSE= the variable is not the variable type specified in the format check function.

Converting a numeric variable to a character variable

```
> dataSepsis$HR_char <- as.character(dataSepsis$HR)
```



New variable name

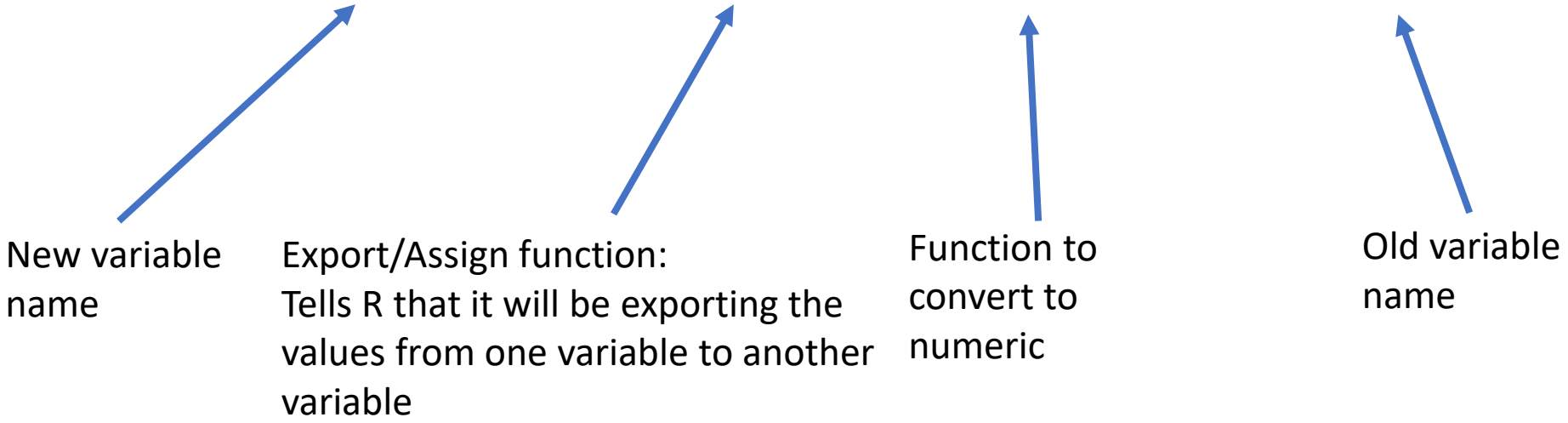
Export/Assign function:
Tells R that it will be exporting the values from one variable to another variable

Function to convert to character

Old variable name

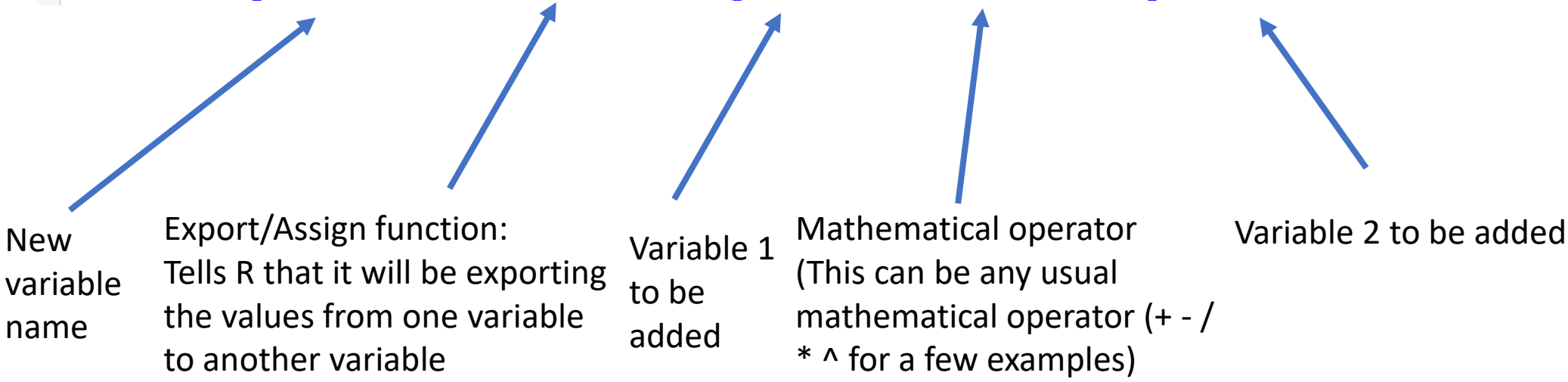
Converting a character variable to a numeric variable

```
> dataSepsis$HR_num <- as.numeric(dataSepsis$HR_char)
```



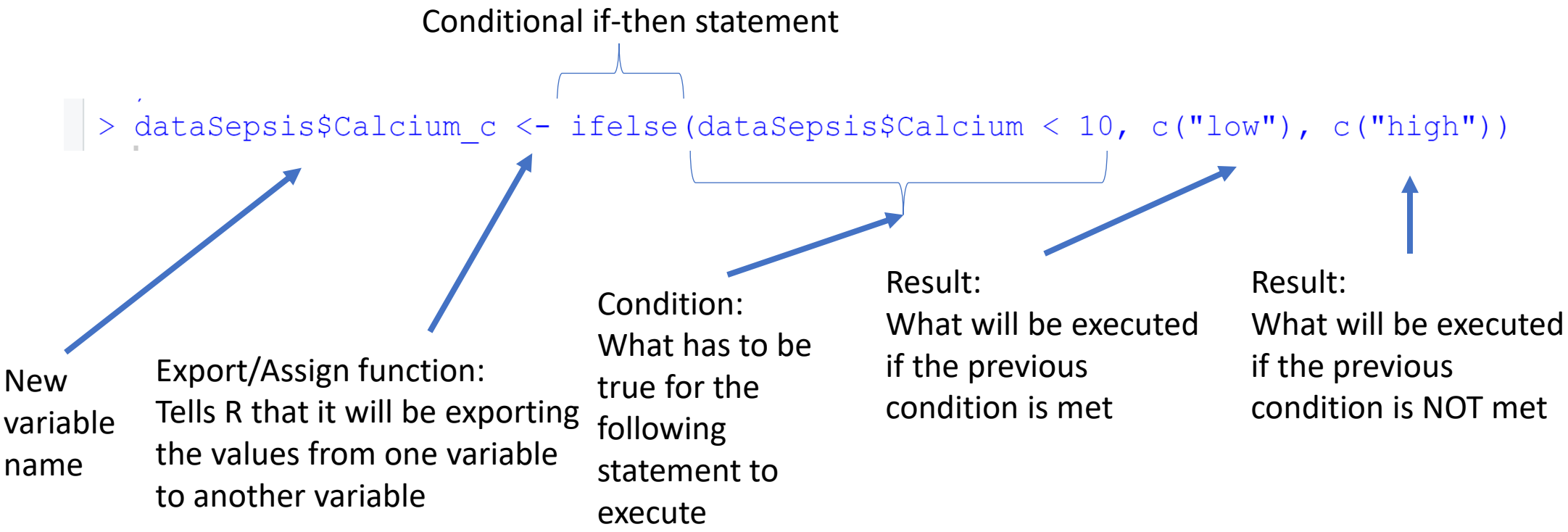
Creating a calculated variable

```
> dataSepsis$Unitsum <- dataSepsis$Unit1 + dataSepsis$Unit2
```



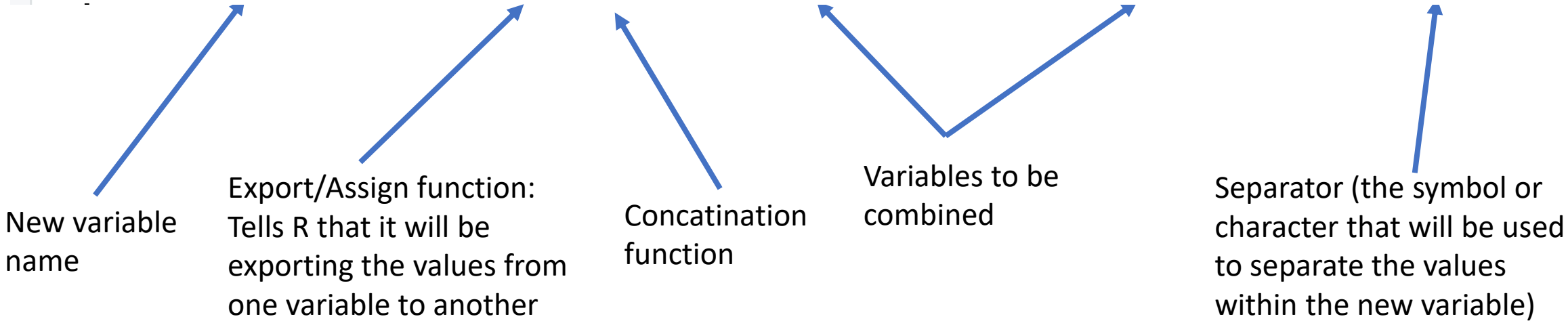
*Calculations can contain multiple variables and multiple operators and functions enclosed in parentheses. See R technical documentation for additional function options that can be used in calculations.

Populating values of a variable using an if-then statement



Combining Multiple Character Variables

```
> dataSepsis$UnitPairID <- paste(dataSepsis$Unit1,dataSepsis$Unit2,sep=",")
```



Taking Only a Portion of a Character String

```
> dataSepsis$SubUnitPairID <- substr(dataSepsis$UnitPairID,1,2)
```